For Whom the Bell Trolls: Shifting Troll Behaviour in the Twitter Brexit Debate

For Whom the Bell Trolls: Shifting 'Troll' Behaviour in the Twitter Brexit Debate

Abstract

Twitter released a list of 2,752 accounts believed connected to state-sponsored Russian operative manipulation of the 2016 American Election. We investigated the behaviour of these accounts in the UK-EU referendum using our longitudinal Twitter dataset. We identified Brexit-related content from 419 of these accounts, totalling 3,485 tweets between 29th August 2015 and 3rd October 2017. While these accounts were primarily designed to resemble American citizens, accounts created in 2016 contained German and Italian locations and terms in user profiles, suggesting targeting of wider international electoral processes. Brexit was one of many targets, likely indicating coordinated repurposing of 'troll' activity over time. We analyse behavioural shifts in account behaviour in relation to external events, introducing a temporal dimension not typical of political Twitter studies. The 'troll' account behaviour altered radically on UK-EU referendum day, shifting from generalised disruptive tweeting to retweeting other troll accounts to amplify their effect.

Introduction

Whether, and to what extent, the activities on Twitter of the Internet Research Agency (IRA) based in St Petersburg (commonly known as a Russian 'troll factory') influenced the outcome of the UK-EU referendum has been the subject of much speculation and of official UK parliamentary inquiry[1]. However, little solid evidence has been available on the nature or extent of this intervention, or on whether it was in fact targeted specifically at the UK's referendum on EU membership.

Trolls are human internet users who attempt to manipulate opinion by spreading rumours, speculation and false information (Mihaylov, Georgiev, and Nakov 2015). Twitter identified 2,752 so-called troll accounts they claim are likely run by the IRA, a Russian company, which was identified as tweeting about the US 2016 elections[2]. We have been collecting tweets on the topic of the UK-EU 'Brexit' referendum since August 2015. Using a sophisticated multiple collection strategy to minimise selection bias (anonymised reference), we have

1 https://www.parliament.uk/business/committees/committees-a-z/commons-select/digital-culture-media-and-sport-committee/inquiries/parliament-2017/fake-news-17-19/

2 https://democrats-intelligence.house.gov/uploadedfiles/exhibit_b.pdf

collected over 70 million Brexit-related tweets. Our findings, using advanced analytical methods including machine-learning to analyse the tweet text and metadata from our derived and aggregated data[3], allow us to provide important insights into the behaviour of these known Twitter 'trolls' over time and in relation to external events. We can also analyse the different types of content that their tweets contain and the implications of this for the detection of different types of human ('troll') verses automated behaviour ('bot') or hybrid automated/human ('cyborg') behaviour deployed in relation to key political events, such as Brexit.

We provide detailed analysis of the behaviour of the IRA-linked accounts identified by Twitter and discuss the strategy and intentions that appear to underlie this behaviour. We find that the scatter-gun disinformation approach observed in the Twitter behaviour of these accounts is consonant with known Kremlin foreign policy strategy, with roots in the Soviet *dezinformatsia* approach (Shultz and Godson, 1986), aimed to disrupt and disorientate foreign regimes and to create a widespread sense of chaos and instability (White, 2016). We find that although targeting the US election was the identifying feature of these accounts (and most indeed masqueraded as bone-fide US citizens), there was evidence of more widespread international agitation, with some of these accounts also generating fake German and Italian user profiles. We do find evidence of Brexit-related activity. In particular, we find evidence of a shift in IRA-related account behaviour on the day of the referendum on UK membership of the EU. Brexit, however, was one of many targets for these accounts, likely indicating a coordinated repurposing of account activity over time as part of a wider disinformation strategy. These accounts, of course, were identified specifically as a consequence of their tweeting about the US 2016 Elections, and there may yet be other unidentified troll accounts that specifically targeted the Brexit referendum.

[3] As a consequence of our ethical procedure, the Twitter Develop Agreement, and following legal advice we are unable to disclose or share usernames, the usernames of retweeted users (unless they are verified users) or any full tweets. We utilise this information in our analysis but it remains confidential. Access to any images or videos contained in these tweets is no longer possible as these have been removed from the Twitter website.

Political Twitter Strategies and the International Disinformation Ecology

The emerging international disinformation ecology and the role played by social-media platforms are topics of increasing academic and policy concern (Derakhshan and Wardle 2017). 'Active-measures', including the funding of outlets to spread disruptive disinformation, are a regular part of the work of the Russian espionage and security agencies, with the specific nature of these activities adapted to target the different points of vulnerability of different foreign states. Disruption, the encouragement of internal divisions and the fomenting of widespread uncertainty, is a common strategic approach to states with strong institutional structures and little affiliation with Russia (Galeotti, 2017).

On-line grassroots movements can be faked, a phenomenon referred to as astroturfing. Essentially this is a form of propaganda activity. Biased and misleading information is coordinated and shared to promote a specific point of view. The exponential evolution of social-media sites allows direct, and fast, communication to and amongst the public. While this approach can be used by governments, political parties and campaign groups (Briant 2015), it can also be used by less formal and more covert forces to spread propaganda. Grassroots social-media activism is thought to have had a significant influence on, amongst other political events, Obama's 2008 US election campaign, the organisation of the 2011 Occupy Wall Street movement (Juris 2012) and Corbyn's 2015 Labour Leadership campaign (Chadwick and Stromer-Galley 2016). The astroturfing effect has been highlighted by Cho et al. (2011) who demonstrate that uncertainty increased and belief in the likelihood that climate change is a real phenomenon decreased amongst those that had been exposed to climate change-denying astroturf websites. Harris et al. (2014) describe the various methods through which the social-media site Twitter is used for distributing astroturf propaganda. They discuss, for example, how a 'Twitter bomb' (increased Twitter activity on a specific subject in a short period of time) was used to promote a false sense of agreement and encourage members of the Chicago City Council to vote against proposed regulation of electronic cigarettes.

The use of automatically generated Twitter content is also commonplace, Bessi and Ferrara (2016) found, for example, that one fifth of the Twitter conversation about the 2016 US Elections was not generated by humans. 'Bot' accounts are set up to automatically retweet and aggregate content from other sources or to create automatically generated text. 'Influence bots' were described in the DARPA (Defence Advanced Research Project Agency) Twitter Bot Challenge as bots designed to influence discussion on social-media sites (Subrahmanian et al. 2016). Social-media companies are not required to fact check information. Catchiness and repeatability can lead to widespread dissemination of content whether it is true or not (Ratkiewicz et al. 2011). Bots are often used as a method for repeating information and making it appear that the information is popular (Ferrara et al. 2016). As bots have become more advanced, they are able to interact with other bots and humans in a conversational type way making them more believable and increasing their social networks (Ferrara et al. 2016). Automatically extracting information from real users and from the wider internet allows the automatic generation of life-like user profile information creating complex and believable 'sock puppet' personas (Ferrara et al. 2016).

Automated accounts can be used to produce large amounts of content on single issues, where many accounts become active and tweet on the same topic at once forming a 'bot-legion' (Chu et al. 2012). Ratkiewicz et al (2011) describe how nine automated fake users tweeted 929 times in 138 minutes in a 2009 Massachusetts election. This type of activity is intended to start a cascade of information spreading with non-automated accounts then also reproducing the content. Messages are also more likely to be believed if they are seen from multiple sources (Ratkiewicz et al. 2011, Del Vicario et al. 2016)). This creates a normalising wallpaper effect – as the information is seen so often it becomes background noise and is assumed to be true.

Cyborgs, have the behaviour patterns of both bots and humans and are at least partially operated by humans (Chu et al. 2012). Cyborgs are harder to detect than bots as they have the

behaviour patterns of both bots and humans. Snap-shot Twitter analyses might easily miscategorise cyborgs as bots, if only one element of their behaviour was captured on any given day.

Astro-turfing may involve the use of bots, cyborgs or sock puppets to emulate the personas of individuals involved in grassroot political movements (Ratkiewicz et al. 2011). Astro-turf-cyborgs commonly combine political information with more general human content. Keller et al (2017) found that cyborgs were used in astroturfing by the South Korean secret service in the 2012 elections. They observed specific behaviour patterns: having many accounts tweet the same tweet at the same time to influence trending topics, having an agent cut and paste roughly the same content into many accounts, and a consistent time pattern for the activity in the accounts. They found that human troll accounts often act in similar and repetitive ways, as these individual trolls are following coordinated central instructions.

The Brexit discussion on Twitter differed from the political debates surrounding general elections, as voters were presented with a binary choice between leaving or remaining in the EU. This choice did not map directly on to the opinions of the mainstream political parties, with the exception of the UKIP party and the 'leave' option. The official campaign groups were newly formed for this referendum and used social-media platforms to mobilise and organise their base from pre-existing and emerging grassroots movements (Usherwood and Wright 2017). The public see digital platforms as a medium for conducting political debate and thereby reshaping the opinions of political parties and in this case referendum campaigns (Chadwick and Stomer-Galley 2016). Howard and Kollanyi (2016), in a study of tweets collected between 5th and the 12th June 2016 in the Brexit referendum, found that bots played a 'small but strategic role in the referendum conversations' but that not all accounts were completely automated. They found that bots played a strategic role in amplifying messages rather than proposing original arguments. They also find that a third of content in their dataset was produced by only one percent of the accounts. They suggest that due to the

amount of content produced, this might indicate bot activity. Bastos and Mercea (2017) found a network of 13,493 bots that tweeted on the Brexit referendum but disappeared after the ballot. They conclude that these accounts were involved in the amplification of human created content.

'Troll' Activity in a Longitudinal Brexit-Related DataSet

Data Sources

On October 31st 2017 Sean Edgett, a legal representative of Twitter, presented evidence to the United States Senate Judiciary Subcommittee on Crime and Terrorism[2]. He provided details of 2,752 accounts that were linked to the IRA. These 2,752 accounts were identified using information obtained by Twitter from third-party sources. The accounts also produced automated content, but approximately 53 percent of their output was produced by humans (referred to as 'trolls', but see our discussion below on the hybrid 'cyborg' nature of these accounts). Twitter studied tweets from 1st September 2016 to 15th November 2016. In written testimony these accounts were described as being 'Russian election-focused efforts'[4]. The troll accounts posed as news outlets, activists, and politically engaged Americans. Edgett's testimony describes the troll behaviour as: contacting prominent individuals through mentions, organising political events and abusive behaviour / harassment.

No equivalent list of Twitter trolls is directly available for the Brexit referendum. As part of a review of 'Fake News', Damian Collins MP, Chair of the UK parliament's Digital, Culture, Media and Sports Select Committee asked that the UK parliament be provided with 'a list of accounts linked to the IRA and any other Russian linked accounts that it [Twitter] has removed and examples of any posts from these accounts that are linked to the United Kingdom'[4]. Twitter responded with six tweets from Russia Today. In the absence of a specific officially-published list detailing accounts from the IRA that were active in the Brexit

[4] http://www.parliament.uk/documents/commons-committees/culture-media-and-sport/171103-Chair-to-Jack-Dorsey-Twitter.pdf

debate, we investigate whether any of the accounts known to be active on the 2016 US Election also produced content related to Brexit. We also analyse the changing nature of their behaviour in relation to the UK's EU referendum.

We have been collecting Twitter data on the Brexit since August 2015. These data allow us to study discussions leading up the referendum and the consequential reaction to the decision of the UK to leave the European Union (anonymised reference). Data were gathered through the Twitter API based on a selection of relevant hashtags chosen by a panel of academic experts. The set of hashtags[5] was updated periodically, to reflect the evolving conversation on Brexit. The dataset currently contains over seventy million tweets.

In our dataset we found 3,485 tweets from the 419 identified troll accounts that were collected between the 29th August 2015 and 3rd October 2017. These tweets contained content about the Brexit vote and related topics that were expected to influence the vote, such as the EU, refugees and migrants. 3,485 is, of course, a tiny proportion of the overall number tweets but it does indicate that some of the same trolls, identified as tweeting about the US elections, were also active in the Brexit debate. We are confident that we will, if anything, have underestimated any Brexit effect.

The user accounts of the trolls identified by Twitter have now been deleted and are not available from Twitter directly. The terms of service of the Twitter Developer Agreement ask that all Tweets are 'deleted within 24 hours after a request to do so by Twitter'[6]. We have an automated method in place that removes all tweets as requested. Therefore, it is entirely possible that some tweets relevant to this study have been deleted, making our findings a conservative estimate of troll activity. Unique archived collection of tweets, such as ours, are

---

5 The hashtags chosen for collection are #eureferendum, #euref, #brexit, #no2eu, #yes2eu, #notoeu, #yestoeu, #betteroffout, #betteroffin, #voteout, #votein, #eureform, #ukineu, #bremain ,#eupoll, #ukreferendum, #ukandeu, #eupol, #imagineeurope, #edeuref, #myimageoftheeu #eu, #referendum, #europe, #ukref, #ref, #migrant, #refugee #strongerin, #leadnotleave, #voteremain, #britainout, '#leaveeu, #voteleave, #beleave, #loveeuropeleaveeu, #greenerin, #britin, #eunegotiation, #eurenegotiation, #grassrootsout, #projectfear, #projectfact, #remaineu, #europeanunion, #brexitfears, #remain, #leave, #takecontrol, #euinorout, #leavechaos, #labourin, #conservatives, #bregret, #brexitvote, #brexitin5words, #labourcoup, #eurefresults, #projectfear, #voteleavelosecontrol, #regrexit, #wearethe48, #scexit, #niineurope, #scotlandineurope, #article50, #scotlandineu

6 https://developer.twitter.com/en/developer-terms/agreement-and-policy

now the only way that academic research into troll activity in the Brexit discussion can be conducted. Our collection strategy means that although we have over 70 million tweets collected we only collect a sample of all tweets on the Brexit debate; there are likely therefore to be additional tweets from trolls on this topic that we will not have collected. While we will have only a proportion of the full content produced by the trolls, as we also have likely the largest longitudinal Twitter data collection on Brexit, this analysis of the captured troll behaviour over time remains of high policy and academic significance.

Analytical Approach

The selection of hashtags used to populate our dataset ensured that we gathered both tweets that related to the Brexit vote directly and also to topics that were expected by experts to influence opinions on Brexit. When researching whether the troll accounts were active in the Brexit discussions we split the data into tweets that were directly about Brexit and those that contained other related topics such as refugees and migration. We annotated tweets we had gathered from the Twitter-identified troll accounts on the basis of whether they were directly about Brexit or not. The annotators were asked to be conservative and only to include tweets in the Brexit set if they were absolutely certain they were directly about the Brexit referendum itself. We found that 1,357 were directly about Brexit, 2,109 were not and 19 were difficult to decide. This gave us 38.94 percent of the tweets that were directly about Brexit. Henceforth we will call these sets 'Brexit Tweets' and 'Related Tweets'. The 19 undecided tweets were excluded from the study. A single coder annotated all tweets and a sub-sample (100) was double coded to validate consistency and measure inter-annotator agreement, thus producing a kappa score of 0.80 (indicative of very high agreement). These tweets contained multilingual content: English, German and Italian. Both annotators were fluent in all of these languages.

Results

Tweets and Retweets

Twitter reported to the Senate committee that, of all the tweets studied from the 1st September 2016 to the 15th November 2016, 1 percent were US election related. Of these 1 percent of election related tweets, 0.74 percent were Russian linked and had been detected by Twitter either as automation or spam. In the report Twitter only considers original tweets; all retweets are excluded[2].

Twitter identified 131,000 tweets from the accounts on the IRA list. Of these, 9 percent were about the American Election (11,790). The total number of tweets identified by Twitter as US Election Related was 189 million. Thus, 0.006 percent of US Election-related content was created by the IRA-related accounts.

In our analysis, we confirm that these accounts were also creating Brexit-related content. In total we found 3,485 tweets from the IRA linked accounts, representing 0.005 percent of the total data we collected. Our 0.005 percent figure includes retweets and drops to 0.002 percent when these are excluded. In our troll dataset 57.59 percent of data are retweets.

The figures indicate that there is a lower level of activity in our set of IRA accounts discussing Brexit. This is perhaps not surprising: (i) accounts created to target the US election might be expected to be less active on the Brexit topic; (ii) Trolls may be more active on certain issues at certain points - the lower level of activity seen here could reflect the longitudinal nature of our dataset. We gathered and analysed data from a period of over two years whilst Twitter presented an analysis of data only from 1st September 2016 until 15th November 2016; and (iii) there are likely other trolls that are more active in the Brexit debate but these are not on the list submitted by Twitter.

Our longitudinal dataset allows us to move beyond a snapshot picture and to track the changes in activity and troll behaviour over time and in response to external events. In particular, we note a change in retweeting behaviour on the 23rd June 2016, the day of the UK-EU referendum vote. On this day we captured 1,059,888 tweets in total. Out of this total, 389 of these tweets were from troll accounts (0.037 percent), over a seven-fold (7.4 percent) increase in the relative number of tweets that came from trolls on the day of the referendum. The vast majority of tweets captured on this day were retweets, something the headline value of 0.74 percent in the US Twitter report would fail to identify. In fact, only eleven tweets were original tweets. If we were to calculate troll activity excluding retweets, we find that on this day 0.001 percent of original data are from trolls, seriously underestimating their potential impact. Indeed, conducting the calculation in this way would indicate a misleading decrease in troll activity rather than an increase. This highlights the marked change in their behaviour on the day of the Brexit referendum vote. Although the trolls were more active, they produced more retweets and less original content. We must therefore consider this when we evaluate the 0.74 percent value given by Twitter.

User Information

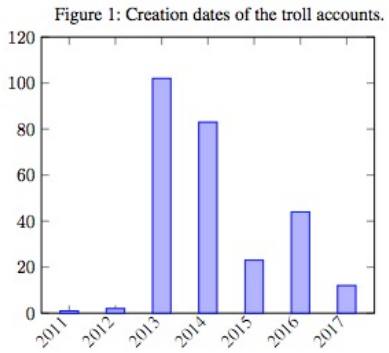
We found tweets from 419 of the 2752 Twitter-identified accounts, with more than one tweet from 66.83 percent of accounts. The Brexit Tweets set we had tweets from 267 accounts, with more than one tweet from 56.68 percent of accounts, indicating that most accounts engaged with the Brexit topic multiple times.

We examined the information from the user metadata as extracted from the troll tweets. This metadata is generated automatically or added by the account holder. The information can change over time and/or be altered by the user. For example, the number of followers is an automated value that can vary, the user profile location field can be added by the account holder and can be changed over time. If the account contains fake information, such as with astro-turf-cyborgs, information that is added by the user can indicate the role and purpose of

the account. An account could be created to look like an American individual with a particular political opinion, expressed through the information added in fields such as location and user description.

Date of account creation is an automatic field and cannot be changed. Most of the accounts were created in 2013 and 2014 (Figure 1) after the 2012 US Election but well before both the 2016 US Election and the Brexit Referendum. More accounts are created in 2016, the year of the US Election (and the Brexit Referendum), than in either 2015 or 2017. Of the forty-four accounts created in 2016 thirty-eight were created after the Brexit referendum but before the US election, all but one of these within a very tight window between 4th and 13th of July.

Figure 1



To further determine whom the accounts were intended to represent, and therefore influence through 'shared' group identities, we analysed the user description field in the tweet metadata. We counted the occurrences of terms by the year of account creation, and we removed very common English and German words. Although a user can change the text in this field at anytime, we did not find any evidence of changes in the data we collected (we gathered data over 2 years and when we had multiple tweets from the same users this field always contained the same content). The full list of terms can be seen in Table 1. Many of the accounts do not have any terms at all in this field. The counts in Table 1 are small but do indicate a pattern. In 2013-2015 the accounts contained description terms that indicate American, conservative, patriotic personas, suggesting that the accounts were designed to

masquerade as US citizens and aimed to influence American events. But in 2016 many of the terms are German, mag (like), glaube (I believe), uern (likely a shortening of äuβern which translates as express). Therefore, the accounts created in 2016 may not, in fact, have been designed to tweet about the US election but something more European-focused.

Table 1: Frequency of terms from the user level description field split by year

2013		2014		2015		2016		2017	
conservative	16	conservative	9	love	5	usa	5	trump	5
blacklivesmatter	15	tcot	6	proud	4	ttip	3	2a	3
love	11	wakeupamerica	5	tcot	4	fr	3	follow	2
tcot	11	patriot	4	family	4	mag	3	mom	2
dont	8	supporter	4	country	4	glaube	2	god	2
pjnet	7	life	4	christian	3	uern	2	starke	1
wakeupamerica	7	pjnet	3	conservative	3	studiere	2	coordinator	1
life	7	dont	3	patriot	2	freizeit	2	broker	1
2a	6	2a	3	youre	2	spiele	2	moment	1

Of the forty-four accounts created in 2016 we found twelve of the accounts created in July 2016 had German language descriptions (of the rest one was in English, one was mixed German and English, two only contained hashtags, and the rest were empty). For comparison, in 2015, sixteen were in English, three in German and four were empty.

The German language use in accounts created after the 2016 Brexit vote suggests that the trolls were using the result of the vote to push a wider disruptive agenda beyond the impact in the UK; perhaps anticipating the German elections in 2017 with Angela Merkel's announcement in November 2016 that she would run for a fourth term as German Chancellor. December 2016 saw Italian elections and the re-vote in the controversial Austrian Presidential election.

We also analysed the user location fields. We found 154 accounts to claim to be based in the USA, sixteen in Europe and three in Russia (94 had no location information). The sixteen

European accounts were made up of seven German accounts, five Italian accounts, three from the UK and one from Belgium. Term counts from the location field are shown in Table 2. The European-based accounts were mainly created in 2016 and later. This location information suggests that most of the accounts on the list submitted by Twitter to US Senate were indeed designed to look like they are from the USA. The agenda they were designed to follow was also related to the USA, but those created in and after 2016 had a different agenda.

Table 2: Frequency of individual terms from the user location field split by year

2013		2014		2015		2016		2017	
usa	51	usa	21	usa	6	deutschland	3	estados	4
states	12	atlanta	10	texas	3	berlin	2	unidos	4
united	12	us	5	germany	2	hessen	1	italia	2
chicago	4	states	3	brussel	1			main	1
us	4	united	3	stlouis	1			lombardia	1
il	4	la	2	tennessee	1			frankfurt	1
ny	3	new	2	states	1			italy	1
baltimore	2	york	2	richmond	1			sicilia	1
ga	2	pittsburgh	1	united	1			itala	1
Atlanta	2	ga	1	wisconsin	1			milano	1

Some of the accounts have many followers and therefore a high potential to reach other Twitter users, thus magnifying the normalising wallpaper effect of their content. Using the maximum number of followers when we had multiple tweets from an account, we found that, of the 267 that had tweeted about Brexit: 122 accounts had more than 1,000 followers; sixteen accounts had over 10,000; and one account had over 100,000. The median number of followers is 875. As we cannot tell from this dataset how many of the trolls follow each other, this high median number should be treated with caution. There is a very slight positive correlation (Pearsons' Rho 0.01) between number of followers and number of captured tweets, this should also be treated with caution as we do not capture all of the tweets from each user.

Tweet Information

The hashtags that are used by the troll accounts indicate the different topics discussed in both the Brexit Tweet and the Related Tweet sets. Table 3 shows the top hashtags in each set. In the Brexit Tweet set the hashtags used are related to Brexit, Britain, and the EU. In the top ten we also find hashtags relating to Chancellor Merkel, #merkel, and #merkelmussbleiben (which translates as #merkelmuststay). The accounts are directly discussing Brexit but also using wider hashtags for example referring to the role of the German Chancellor Merkel, underlining the wider European context of the Brexit debate.

In the Related Tweet set we can see that the trolls use hashtags about the EU, #eu; about refugees, #refugeeswelcome, #flchtlinge (which translates as #refugee) and #refugee. We see mentions of the German Chancellor, #merkel and the President of Turkey #erdogan, we also see reference to Germany, #deutschland and Turkey #trkei. We also see that tweets that were classified as not directly about Brexit are still being tagged with the #brexit hashtag.

The way that the hashtags are being used in the wider Brexit related set suggests that the trolls have an agenda that related to Germany and Turkey and were using the Brexit topic to push this agenda and create a sympathetic audience on the controversial wider issue of migration. Elections were held in Germany on 24th September 2017, and in Turkey there was a constitutional referendum held on the 16th April 2017.

Table 3: Hashtags frequency across the data specifically on Brexit (Brexit) and those on Brexit related tweets (Related)

Brexit		Related	
#brexit	825	#eu	1206
#britaininout	378	#merkel	286
#euref	364	#refugeeswelcome	281
#brexitornot	211	#flchtlinge	199
#goodbyeuk	188	#erdogan	158

#brexitinout	186	#europe	153
#remainineu	168	#deutschland	146
#eu	158	#trkei	128
#merkelmussbleiben	125	#brexit	113
#merkel	111	#refugee	80

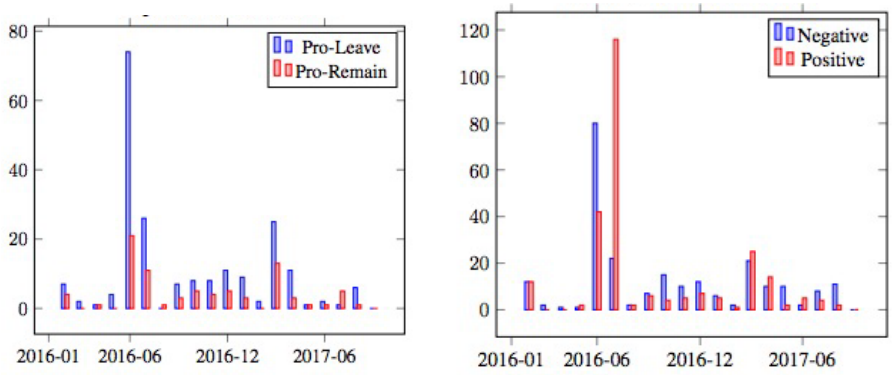
Sentiment and Stance

We annotated the 1,357 Brexit Tweets for both stance and sentiment. The annotator was asked to rate the stance of the tweets as either pro-leaving or pro-remaining in the EU or neutral/neither. For sentiment the tweets were annotated as containing positive, negative or neutral sentiment. The majority of tweets were both neutral in stance (78.78 percent) and sentiment (64.41 percent), although this may be a consequence of the lack of available image or video context. Overall the tweets had a stronger pro-leave stance (14.96 percent) than pro-remain (6.26 percent). The split of sentiment was fairly equal with a positive sentiment (18.35 percent) being very slightly higher than negative sentiment (17.24 percent).

The pro-leave stance was consistently higher throughout the time period (Figure 2a). The sentiment scores do change over time (Figure 2b). In particular there was a spike of positive sentiment tweets on the 21st July 2016. This was a spike in volume that occurred on the day that UK Prime Minister May met German Chancellor Merkel (see next section). A high percentage of these (73.44percent) were original content not retweets and the tweets were in German. The content driving this change in sentiment direction revolves around Chancellor Merkel, describing her as a strong person that will handle the Brexit issue well. The trolls discuss new possibilities and options after Brexit and that Frankfurt will be soon in a stronger position. A few trolls also talk about the EU accession of Turkey and that Merkel does not want any negotiations if Turkey re-introduces the death penalty.

Figure 2 a) Stance of Tweets and b) Sentiment of Tweets

a) b)

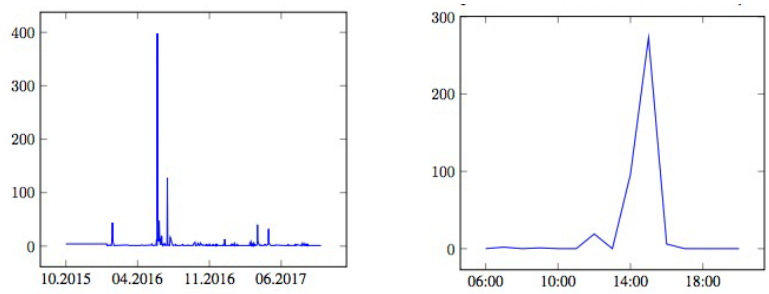


External Events

Overall there were very few tweets per day from the troll accounts. The number of tweets per day follows a similar pattern in both the Brexit Tweet and the Related Tweet sets. In both cases, we find these patterns to be closely related to external events. We can see in Figure 3 that there are several spikes of activity for the Brexit Tweets dataset.

Figure 3. Number of Tweets over time a) by day across the full dataset b) by hour on the 23rd June 2016

a) **b)**



In Table 4 we show the dates, volumes, whether the tweet was original content or a retweet and what was happening in the news that may have triggered the content. We found that the spikes in content production are related to particular events such as the referendum vote itself or the UK Prime Minister May meeting German Chancellor Merkel. There is a difference in troll behaviour on two particular dates that exhibit higher tweet volumes. On the 23rd June 2016, the day of the UK's EU referendum, there is a large increase in volume of tweets

produced that consists almost entirely retweets (97.73 percent). In contrast, on the 21st July 2016 there was a considerably larger proportion of original content produced: only 26.56 percent on this day are retweets. There is also a conspicuous spike on the 19th February 2016. On a closer inspection of the data from this day, we found that all the tweets come from a single troll account.

Table 4: Further information about the days that had the highest frequency of tweets

Date	Tweets	percent RT	What happened on that day?
19/02/2016	44	100.00	Cameron at EU summit
23/06/2016	398	97.74	UK-EU referendum
24/06/2016	51	49.02	Day after UK-EU referendum
28/06/2016	47	70.21	Cameron to meet EU leaders
21/07/2016	128	26.56	May meets Merkel
29/03/2017	40	45.00	UK triggers Article 50 process
29/04/2017	32	100.00	EU Council Guidelines for Brexit Negotiations issued

The Day of the Brexit Referendum

The largest number of troll tweets was collected on the day of the referendum, the 23rd of June 2016. The difference in content on this date is shown in Table 5. We collected 400 in total of which 398 were directly about Brexit. It is difficult to compare troll behaviour with the behaviour of an average Twitter user as the median number of tweets from a user in our set is 1 (there are many tweets from a few users but the majority only tweets once), but this overall surge in number of tweets contrasts starkly with non-troll account behaviour which saw a surge on the day after Brexit when the referendum result was known. On referendum day, of the 398 Brexit troll tweets, only nine (and eleven tweets out of the 400 total) consisted of original content, the rest were retweets. 97.73 percent of tweets on the day of the referendum were retweets. The trolls were therefore focused entirely on Brexit and shifted to retweeting rather than producing original content.

Out of the 387 retweets 279 were retweets of other trolls from the list issued to Senate by Twitter (72.10 percent). These tweets originate from only 11 troll accounts, and 186 were retweets originating from a single troll account. These 186 tweets exhibit a very similar homogeneity in style, content and format (text in italics altered from original):

@USER #brexitornot #britaininout #brexitinout #euref https://t.co/*VARIOUS*

Table 5: The most frequent hashtags from the brexit dataset on the 23rd June 2016, the day of the referendum and on all other days

All other days		23rd June 2016	
#brexit	787	#britaininout	378
#eu	148	#euref	354
#merkelmussbleiben	125	#brexitornot	211
#merkel	110	#goodbyeuk	187
#euco	89	#brexitinout	186
#ukineu	64	#remainineu	168
#may	60	#brexit	38
#uk	33	#reasonstoleaveeu	14
#article50	31	#eu	10
#girlstalkselfies	18	#uk	8

In Figure 3b we see the frequency of tweets grouped by hour across the day of the referendum. The vast majority of tweets were sent between 2pm and 4pm in a highly concentrated effort. There were no tweets after 4pm although the polls did not close until 10pm. The nine tweets consisting of original content were tweeted early in the day, and two original tweets were tweeted after 2pm (at 2pm and 2.45pm).

Amplification Behaviour

A social-media amplifier is defined as a user that shares ideas and opinions (Tinati et al. 2012). Amplification is a key element of astro-turfing for grass-root twitter mobilisation. In this context we will use the term amplifier to classify an account that, as far as we can see from the data we have collected, only ever retweets.

We have a large number of trolls who are amplifiers in our set. To judge if their behaviour changed on the referendum polling day we analysed troll accounts which sent tweets on both the 23rd June and other days. In the Brexit Tweets set we have tweets from 248 troll accounts, of which 38 of them were active on the 23rd June 2016. Of those accounts, 19 (50 percent) also appear on other days. The other 19 troll accounts may only have tweeted about Brexit on the 23rd June or it could just mean we did not catch them in our dataset.

Only nine tweets were not retweets on the 23rd June. Those original content tweets all came from accounts that tweeted on other days as well. In this dataset the accounts that only tweeted on the 23rd June 2016 were amplifiers. Thirteen of the accounts acted as amplifiers on the 23rd June and twelve in the wider time period. As we do not have all of the tweets produced by all trolls we cannot establish a definitive pattern but this suggests that, while some accounts may simply be amplifiers, content producers can also switch their behaviour to amplification if required. These trolls may be more likely to be one or another but these behaviour patterns can change. On the day of the referendum vote we found that all of the IRA troll accounts were more likely to be involved in amplification behaviour.

Discussion

The IRA accounts identified by Twitter as specifically tweeting about the US 2016 elections, were also active in the Brexit debate. A limitation of this study is that it is likely we do not have all of the content produced on this topic and, as the content is no longer available from Twitter, there may be more content that we have not analysed here. We also can not speculate

what these accounts were doing when they are not tweeting about either the US election or Brexit as this data is not available.

We have observed that these trolls do tweet about the Brexit but the volume is small. Given that this is a 'low cost behaviour' we might have expected to see more activity if aim of these accounts was to influence the referendum. We only find 3,485 tweets in a dataset of over 70 million. On the other hand, it is perhaps surprising that we find content on Brexit from these accounts at all, as they were accounts submitted by Twitter to the US Senate committee that were thought to be attempting to influence the US Elections. Why then do they tweet about Brexit at all if this is not their main agenda? Consistent with known Russian disinformation strategies and 'active-measures' we suggest that Brexit as a controversial issue with wider implications in relation to, for example, international free-trade deals and for EU-wide stability, provided a suitable topic for generalised disruptive tweeting. This also suggests that there are likely other accounts specifically designed by the IRA to look like Brexit grassroots activists that have not yet been released by Twitter.

The tweets that we collected from troll accounts were slightly more likely to be pro-leave than pro-remain. These accounts were not, however, designed to look like either pro-leave or pro-remain grassroots individuals; this is unsurprising as it appears that influencing the Brexit referendum was not a priority for these troll accounts. Given the source of the list and the characteristics we have uncovered, these were designed to be active in the US and perhaps latterly in the German elections. This is supported by the dates of creation of the accounts and the personas used; 2013 and 2014 accounts were designed to look like American activists and 2016 were designed to look German.

The Brexit tweets analysed here may, of course, simply constitute background noise, designed to make the American and latterly German sock-puppet accounts look either more human or more politically aware. It is clear, as some accounts were created as far back as

2013, that the IRA is was playing a long game rather than seeking always to directly influence voting behaviour. Spikes in data production, were closely related to external events. We see, for example, on the 21st July there was a high level of positive original content produced that was related to Merkel and Brexit, certainly pointing towards the use of Brexit in wider political context, not only in relation to internal UK Brexit dynamics.

The strategy employed by these accounts clearly shifted on the day of the referendum where we observe an apparent shift to amplification behaviour on the Brexit issue. The change in behaviour on this specific date, the day of the referendum vote, indicates a possible attempt to directly influence public behaviour. The longitudinal nature of our data collection technique allowed us to investigate these behavioural changes and adaptations. Many retweets that contain very similar content were tweeted over a short time frame. The IRA-identified trolls began almost exclusively to retweet other trolls on Brexit issues on referendum day, with virtual abandonment of the generation of original content. This change in behaviour indicates a clear change in strategy from the IRA, this may of course be coincidental but, the date of change makes this seem unlikely, especially as the account behaviour returns to normal after this event.

This observed change in behaviour indicates a possible method for cyborg identification. The accounts we tracked combined apparently automated 'bot' behaviour and human troll activity, suggesting that these are in fact cyborg accounts. Automated behaviour can be automatically detected as opposed to human 'troll' behaviour, which is more difficult to detect. If we had studied the data from the 23rd June 2016 in isolation then the cyborg troll accounts would simply have resembled bot accounts, and might easily have been mis-classified. To successfully identify the cyborg accounts it was necessary to systematically look for changes in behaviour over time. If we can identify accounts that combine behaviours this would enable us to develop better methods to identify these harder to spot cyborg accounts.

It is likely that the referendum day tweets were produced automatically or exhibit the cut and paste behaviour seen in the South Korean Election (Keller, 2017). It appears that the operators of the American and German personae had been instructed to tweet on the Brexit topic en-masse, suggesting that sock-puppet accounts were being re-purposed to target international electoral developments - shifting their behaviour and the target of activity as required or as instructed when significant or potentially disruptive events emerge. In which case, overlapping patterns of troll activity are likely to be observed and any forthcoming lists of Brexit trolls, while a welcome starting point, should be treated with caution.

Conclusion

All 2,752 of the accounts identified by Twitter as having IRA-links have since been suspended and the information posted by them is therefore no longer available through Twitter, making our archive a unique source of insight into so-called troll behaviour on the Brexit issue. Despite the absence of an equivalent list of 'trolls' that sought to target the Brexit referendum specifically, there is clear evidence that the IRA-directed activity altered dramatically on the day of the referendum to target this event. Our longitudinal approach suggests that these accounts exhibited behaviour consistent with hybrid human/automated activity, raising questions about potential mis-classification of bot/troll activity in short-term studies. Much of the observed activity in our dataset post-dated the referendum and their activity was not directed solely at the internal UK Brexit process. Rather it appears that this was a part of a more widespread, centrally coordinated IRA effort to influence international electoral processes, with troll activity temporarily diverted and repurposed to the Brexit case. This is consistent with known Kremlin disinformation approaches, and 'active-measures', utilising controversial topics to escalates underlying uncertainty and to create a sense of mistrust, instability and insecurity in foreign regimes.

Acknowledgements

This work was conducted as part of the UK in a Changing Europe programme and was funded by the Economic and Social Research Council grant no. ES/N003985/1. We would like to thank the three anonymous reviewers for constructive feedback.

References

Bastos, M. T., and Mercea, D. 2017. The brexit botnet and user-generated hyperpartisan news. Social Science Computer Review

Bessi, A., and Ferrara, E. 2016. Social bots distort the 2016 us presidential election online discussion. First Monday 21(11).

Briant, E.L., 2015. Propaganda and counter-terrorism: Strategies for global change.

Chadwick, A. and Stromer-Galley, J., 2016. Digital media, power, and democracy in parties and election campaigns: Party decline or party renewal?

Cho, C.H., Martens, M.L., Kim, H. and Rodrigue, M., 2011. Astroturfing global warming: It isn't always greener on the other side of the fence. *Journal of business ethics*, *104*(4).

Chu, Z.; Gianvecchio, S.; Wang, H.; and Jajodia, S. 2012a. Detecting automation of twitter accounts: Are you a human, bot, or cyborg? IEEE Transactions on Dependable and Secure Computing 9(6):

Davis, C. A.; Varol, O.; Ferrara, E.; Flammini, A.; and Menczer, F. 2016. Botornot: A system to evaluate social bots. In Proceedings of the 25th ICCWWW

Del Vicario, M.; Bessi, A.; Zollo, F.; Petroni, F.; Scala, A.; Caldarelli, G.; Stanley, H. E.; and Quattrociocchi, W. 2016. The spreading of misinformation online. Proceedings of the National Academy of Sciences 113(3)

Wardle, C. and Derakhshan, H., 2017. Information Disorder: Toward an interdisciplinary framework for research and policymaking. *Council of Europe report, DGI (2017)*, *9*.

Ferrara, E.; Varol, O.; Davis, C.; Menczer, F.; and Flammini, A. 2016. The rise of social bots. Communications of the ACM 59(7).

Galeotti, M., 2017. Controlling Chaos: How Russia Manages Its Political War in Europe. *Policy Brief, European Council of Foreign Relations*.

Harris, J.K., Moreland-Russell, S., Choucair, B., Mansour, R., Staub, M. and Simmons, K., 2014. Tweeting for and against public health policy: response to the Chicago Department of Public Health's electronic cigarette Twitter campaign. *Journal of medical Internet research*, *16*(10).

Howard, P. N., and Kollanyi, B. 2016. Bots, #strongerin, and #brexit: Computational propaganda during the uk-eu referendum.

Juris, J.S., 2012. Reflections on #OccupyEverywhere: Social-media, public space, and emerging logics of aggregation. *American Ethnologist*, *39*(2)

Keller, F. B.; Schoch, D.; Stier, S.; and Yang, J. 2017. How to manipulate social-media: Analyzing political astroturfing using ground truth data from south korea. ICWSM

Mihaylov, T.; Georgiev, G.; and Nakov, P. 2015. Finding opinion manipulation trolls in news community forums. In CoNLL

Ratkiewicz, J.; Conover, M.; Meiss, M. R.; Gonc¸alves, B.; Flammini, A.; and Menczer, F. 2011. Detecting and tracking political abuse in social-media. ICWSM

Shultz, R.; and Godson, R.S. 1986 Dezinformatsia The Strategy of Soviet Disinformation. London: Penguin.

Subrahmanian, V.; Azaria, A.; Durst, S.; Kagan, V.; Galstyan, A.; Lerman, K.; Zhu, L.; Ferrara, E.; Flammini, A.; and Menczer, F. 2016. The darpa twitter bot challenge. Computer 49(6)

Tinati, R.; Carr, L.; Hall, W.; and Bentwood, J. 2012. Identifying communicator roles in twitter. In Proceedings of the 21st ICWWW

Usherwood, S. and Wright, K.A., 2017. Sticks and stones: Comparing Twitter campaigning strategies in the European Union referendum. *The British Journal of Politics and International Relations*, *19*(2).

White, J., 2016. Dismiss, Distort, Distract, and Dismay: Continuity and Change in Russian Disinformation. *Institute for European Studies Policy Brief 13*.